Journal of Biomedical Informatics 45 (2012) 1185–1190

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Workflow mining and outlier detection from clinical activity logs

L. Bouarfa *, J. Dankelman

Department of Biomechanical Engineering, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Article history:
Received 12 February 2012
Accepted 9 August 2012
Available online 19 August 2012

Keywords:
Medical environment
Surgical workflow
Surgical consensus
Outlier detection
Laparoscopic surgery

ABSTRACT

Purpose: The purpose of this study is twofold: (1) to derive a workflow consensus from multiple clinical activity logs and (2) to detect workflow outliers automatically and without prior knowledge from experts.
Methods: Workflow mining is used in this paper to derive consensus workflow from multiple surgical activity logs using tree-guided multiple sequence alignment. To detect outliers, a global pair-wise sequence alignment (Needleman–Wunsch) algorithm is used. The proposed method is validated for Laparoscopic Cholecystectomy (LAPCHOL).
Results: An activity log is directly derived for each LAPCHOL surgery from laparoscopic video using an already developed instrument tracking tool. We showed that a generic consensus can be derived from surgical activity logs using multi-alignment. In total 26 surgery logs are used to derive the consensus for laparoscopic cholecystectomy. The derived consensus conforms to the main steps of laparoscopic cholecystectomy as described in best practices. Using global pair-wise alignment, we showed that outliers can be detected from surgeries using the consensus and the surgical activity log.
Conclusion: Alignment techniques can be used to derive consensus and to detect outliers from clinical activity logs. Detecting outliers particularly in surgery is a main step to automatically mine and analyse the underlying cause of these outliers and improve surgical practices.

1. Introduction

Today's operating rooms (ORs), post anaesthesia care units (PACUs), intensive care units (ICUs) and hospital wards generate vast quantities of workflow data. Sophisticated monitoring equipment performs continuous, high-frequency measurements of multiple physiological parameters. Timely and context-sensitive analysis of these physiological parameters is required in order to provide effective decision support. Because much of this data remains unused, its exploitation could enrich and extend medical/hospital knowledge by identifying the variables that are the most important in increasing throughput (i.e. efficiency) and predicting adverse (i.e. safety) or favourable (i.e. effectiveness) outcomes.

Workflow mining is a technique that aims to improve the workflow modelling process by providing tools for discovering, comparing, and conformance checking of workflow process models [1]. Conformance checking is crucial in the medical domain to detect outliers (e.g. adverse events) from medical protocols; real-time detection of outliers can be used for early error prediction and early warnings for the surgical team to avoid surgical errors. Moreover, workflow discovery and comparison are tools that can be used to enrich and extend the medical protocols using used surgical practice. As such, workflow mining is an essential tool for the continuous modelling of surgical workflow.

1.1. Medical workflows are unusual

Workflow variability is inherent to medical environments [2,3]; it is caused by uncertainty in patient anatomy, unexpected complications, cognition and situational awareness of medical practitioners. Moreover, unlike organisational workflow, medical workflow lacks direct human–computer interaction, which makes automatic activity logging a necessary step before modelling [4]. For this reason, workflow activity monitoring in medical environments is non-trivial. Implementing a workflow analysis in such environments requires an approach for dealing with the variability and monitoring of medical activities [5,6].

A consensus workflow for a specific medical treatment can be constructed using expert opinion; however this can require a lengthy debate, with no guarantee of reaching a final consensus. Another option is to derive the consensus workflow from multiple activity logs extracted from the medical environment without any prior knowledge from experts. This has the advantage that it would reflect real medical practice and would not require expert opinion. The remainder of this section presents an approach to deriving a consensus using individual activity logs.

1.2. Medical data remains unused

In medical environments (ORs, ICUs, PACUs and hospital wards) it is routine practice to measure the patient's blood pressure, heart rate, cardiac rhythm, expired CO_2 and temperature. In the OR,

* Corresponding author. Address: Department of Biomechanical Engineering, Delft University of Technology, Mekelweg 2, 2628CD Delft, The Netherlands.
E-mail address: loubna.bouarfa@gmail.com (L. Bouarfa).

1532-0464/$ - see front matter
http://dx.doi.org/10.1016/j.jbi.2012.08.003

laparoscopic video can be used to track surgical tasks during the intervention. Following the surgery patients are transferred to the PACU or the ICU units where other information can be recorded.

Although much of this data is used for direct medical analysis by medical practitioners, it is not used to enrich and extend medical knowledge (e.g. for workflow mining or adverse event prediction). To allow (statistical) analysis and the comparison of the data, it needs to be structured and synchronised in a single manageable file format. This file is known as the activity log. Constructing the overall activity log is a vital step to allow workflow mining.

1.3. Generating a log file

The first step towards automatic workflow mining in healthcare is the generation of a representative activity log. However, activity logs are not yet extracted and recorded in medical environments. By using pattern recognition techniques on the different data sources (e.g. video, monitored activities, etc.), activities can be identified. Each entry in an activity log represents an activity, for which the start and end times are logged. The entire clinical workflow is represented by the sequence of activities (i.e. entries) contained in the activity log.

1.4. Sequence alignment for consensus and outlier detection

Sequence alignment has been used [7] in bioinformatics since 1970 as a way of finding overlapping or similar sequences of DNA, RNA, or proteins and identify important relationships. It deals with the problem of grouping together sets of sequences to identify areas of similarity between the sequences. Gaps are inserted between the elements of the sequence to align similar characters in successive columns. Consensus derivation from aligned sequences was introduced in 1987 by Gribskov et al. [8]. Taking its inspiration from biological sequence alignment, Jagadeesh Chandra Bose et al. [2] proposed applying this technique for process mining. The processes described in [2] are derived from machine use-logs. In this paper we propose to use an adjusted alignment technique for clinical workflows. The challenges are to accommodate the high variability in the execution of medical procedures and construct the activity logs.

Outlier detection is another challenging problem that aims to identify objects in a dataset whose behaviour is different from the other objects in the same dataset [9]. Outliers represent unusual situations. When detected, outliers often give domain experts the opportunity to extract useful knowledge. In situations where the data is a time series, we need to detect outliers in a sequence of information. In this paper we propose to apply the classical Needleman–Wunsch algorithm [7] to detect outliers that deviate from the consensus in an activity log.

1.5. On workflow mining for laparoscopic surgery

In laparoscopic surgery, the surgeon inserts long, thin instruments into the abdominal cavity through small incisions while an assistant holds a laparoscopic camera. Because of the small incisions, this is considered to be less invasive and traumatising for the patient than open surgery. Laparoscopic video is the main input of workflow activities that can be analysed for workflow mining.

In our previous work, we considered Laparoscopic Cholecystectomy (LAPCHOL) procedures for probabilistic workflow modelling [4]. LAPCHOL is a highly standardised surgical procedure in which a patient's gallbladder is removed because it is inflamed. The approach presented in [4] deals with the variability of LAPCHOL workflows using a probabilistic and prior-knowledge Markov-based approach to detect the various steps in the surgical workflow from instrument activity logs. The high standardisation of the LAPCHOL procedure, together with the results achieved with the probabilistic approach, challenge us to test a framework for generating a consensus and detecting outliers using workflow mining techniques without any prior knowledge.

This paper presents a framework that enables workflow mining for offline utilisation in medical environments intended for the quantitative mining of medical workflows. It proposes a workflow mining framework that allows for the variability of real-time procedures in medical environments. The framework can be applied in any clinical environment with activity logs similar to the ones described in this paper. The framework is evaluated using laparoscopic cholecystectomy. The general framework is presented in Sections 2 and 3 shows how to apply this framework for laparoscopy. Section 4 concludes this paper and highlights the future direction of work.

2. Method

This section describes a framework that allows for the construction of a consensus and the detection of outliers based on this consensus. Fig. 1 illustrates the framework of the workflow mining system, in the following steps:

- The first step is to generate activity logs from the main data sources representing the clinical workflow in a specific medical environment. Section 2.1 describes how to generate an activity log from laparoscopic surgery using laparoscopic video as the main data source.
- The second step is to make use of a large number of activity logs for the same type of treatment and for the same environment to derive a general consensus. This step is performed offline and requires a large number of activity logs. A multi-alignment algorithm is used to generate the consensus as described in Section 2.2.
- The final step is to detect outliers, also known as anomalies, for a specific medical workflow. This step takes place at the end of the workflow (e.g. end of the treatment), by comparing the activity log generated to the consensus derived in the previous step. This is also described in Section 2.2.

2.1. Generate surgical activity log from laparoscopic video

In laparoscopic surgery, laparoscopic video is the main input of workflow activities that can be analysed to arrive at a consensus. In this section, laparoscopic video is used to generate the activity log.

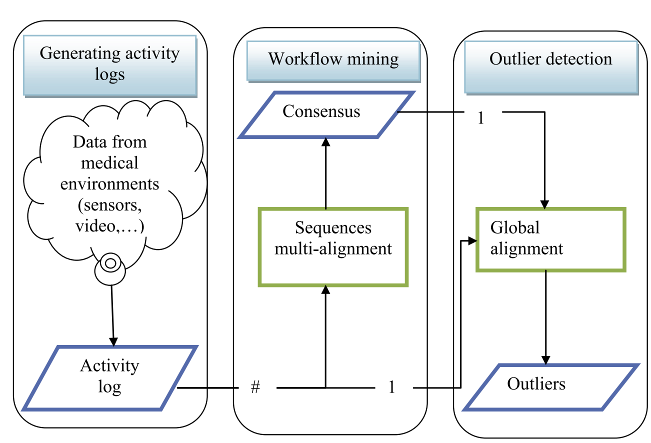

Fig. 1. General framework for workflow mining and outlier detection in surgery.

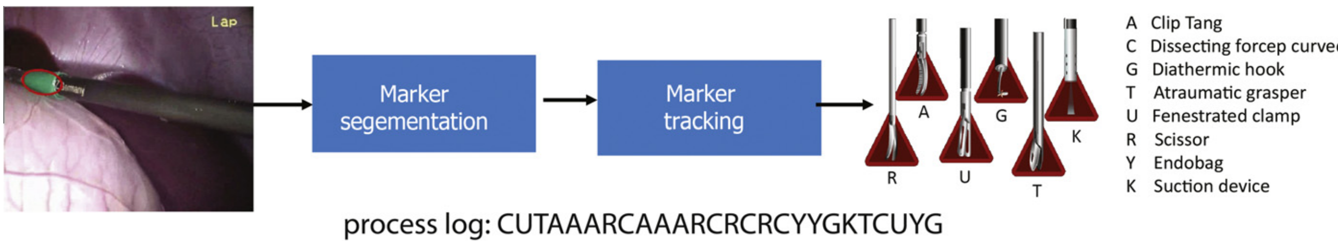

Fig. 2. Instrument tracking tool to produce a surgical activity log from laparoscopic video.

Table 1
An example of an activity log extract from a Lapchol surgery.

Instrument symbol	Start time	End time
C	08:09:27:20	08:10:15:40
U	08:11:07:12	08:11:07:12
T	.	.
A	.	.
A	.	.
A	.	.
R	.	.
C	.	.
A	.	.
A	.	.
A	.	.
R	.	.
C	.	.
R	.	.
C	.	.
R	.	.
C	.	.
Y	.	.
Y	.	.
G	.	.
K	.	.

For this reason, a previously developed tool [10] is used to detect and track single and multiple instruments in laparoscopic video using biocompatible colour markers. Fig. 2 illustrated the tracking algorithm. Initially, the colour markers placed on the instruments are segmented from the background using colour information. Afterwards, the markers are tracked in the segmented regions to derive their trajectories and produce the activity log.

For easier segmentation, the colour markers are selected using a HSV (hue, saturation and value) colour map such that they are positioned with maximum distance relative to each other in the hue space. To build a marker-colour histogram in hue-saturation colour space different images are captured under various illumination conditions and from different parts of the laparoscopic video. Finally, a binary mask is created by thresholding the pixel probabilities. Morphological closing (dilation) is performed on the resulting mask to fill possible gaps in the masks and extend the marker search region for robustness.

To track the instrument over time, blob detection is performed on the mask to group foreground pixels. Afterwards, the markers that are positioned inside these groups are tracked using the OpenCV implementation of a Continuously Adaptive Mean Shift (CAMShift) algorithm [11]. A separate tracker for each marker is used together with its colour model. Finally, a Kalman filter [12] is used together with the CAMShift tracker in order to cope with situations in which the markers are not visible – they could be lost due to image noise or occluded by other instruments or abdominal tissue. In this final step, an ellipse is fitted to the tracked markers. The centre of the ellipse is used as the position of the marker. The output of the tracking tool is an activity log file with the symbol of the instrument used at each entry and the duration of its use, as can be seen in Table 1. In this paper, we discard time information from the analysis (this can always be retrieved afterwards) and focus on the activity log with the activity symbols as presented in the first column of Table 1.

2.2. Consensus derivation and outlier detection strategy

Fig. 3 illustrates the proposed method and the corresponding Matlab code of the framework. It shows how to derive the consensus from surgical activity logs using multiple sequences and how to detect outliers for single surgical activity logs using global alignment. It is important to note that to derive the consensus we used Matlab alignment tools available for aligning ambiguous nucleotide sequences.

To generate the sequence alignment, a distance-matrix was constructed by calculating pair-wise distances between all activity logs. This matrix was used to build a guide tree using a neighbour joining algorithm. The matrix is used by the guide tree to choose the order of pairwise alignments, starting by aligning the closest relatives (neighbouring leaves) and adding the more distant and diverged sequences last. In the next step, the sequences were progressively aligned using the guide tree that defined the order in which the sequences were aligned in the alignment step.

From the multiply aligned sequences, a matrix was calculated with the frequency of each symbol for every column in the multiple alignment, including gaps. The consensus character corresponds to the most frequent character of the sequence profile. When more than one character had a high frequency in the profile sequence, we chose the character that was indicated by the surgeon to be more robust in describing the workflow (in our case of Lapchol procedures the priority ranking is the following: 'ARGKCYTU-'). This is a provisional solution with minor prior knowledge until enough labelled data becomes available to refine the classification of outliers and consensus based on an objective cost-criterion.

Finally, the outliers were calculated by aligning the consensus values and a specific activity log using the Needleman–Wunsch global alignment algorithm. Because the Needleman–Wunsch implementation of Matlab assumes a nucleotide sequence alphabet (NT), the mismatched characters caused by the (NT-assumption) were corrected in a post-processing step. It should be noted that to detect outliers, we used an identity substitution matrix as a scoring matrix, which means that each instrument was as similar as possible to itself and could not be replaced by any other instrument. In reality, instruments can be used for dual functionalities, but more data needs to be collected to define a domain-specific scoring matrix for Lapchol surgeries.

3. Results

3.1. Consensus

Fig. 4 illustrates the results of the alignment algorithm described in steps 5 and 6 in Fig. 3. The derived consensus (CUTA-

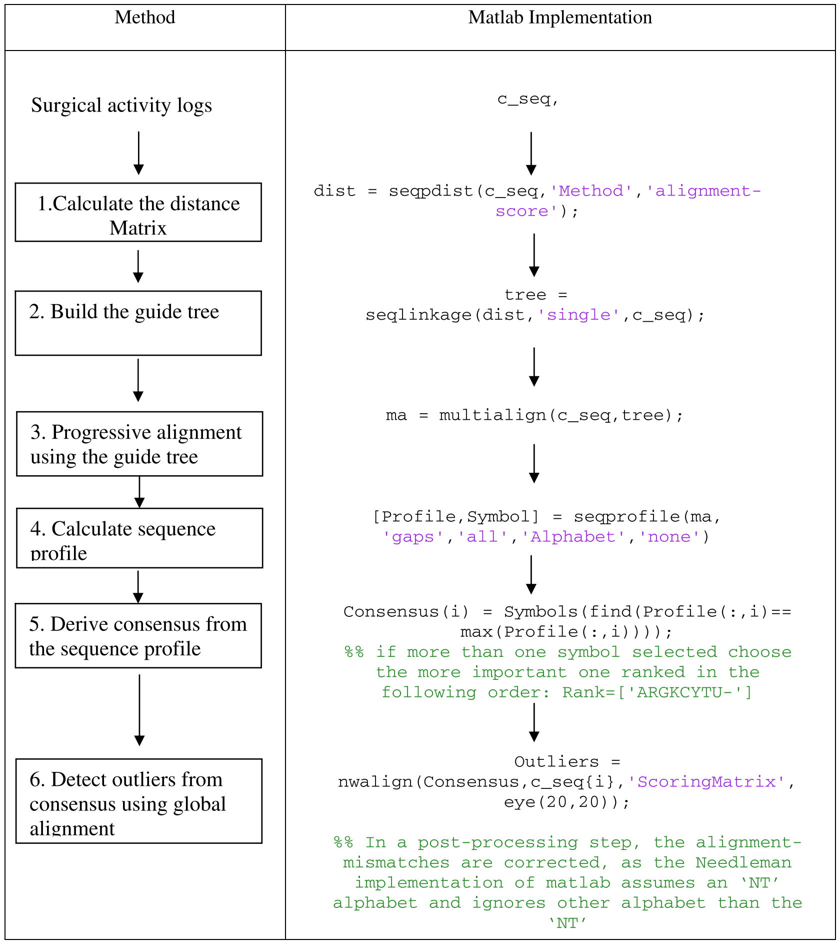

Fig. 3. Consensus derivation and outlier detection steps and corresponding Matlab code, extra comments are given in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

AARCAAARCGKCUYG) represents the main steps of LAPCHOL workflow. The primary step of this surgical procedure is to dissect the area which includes the bile duct and the cystic artery (i.e. Calot's triangle). The main instruments used in this step are represented by the characters C, U and T. When both structures are clearly visible, each of them is clipped in three places, (AAA), and then cut and dissected between the clips with laparoscopic scissors (RC). The following step is the dissection of the gallbladder (C). In minimally invasive surgery, this is done by touching the areas between gallbladder and liver and applying cutting current. The suction device (K) is used continuously, during the last steps, to clean the dissection area. To remove the dissected gallbladder, a salvage bag is

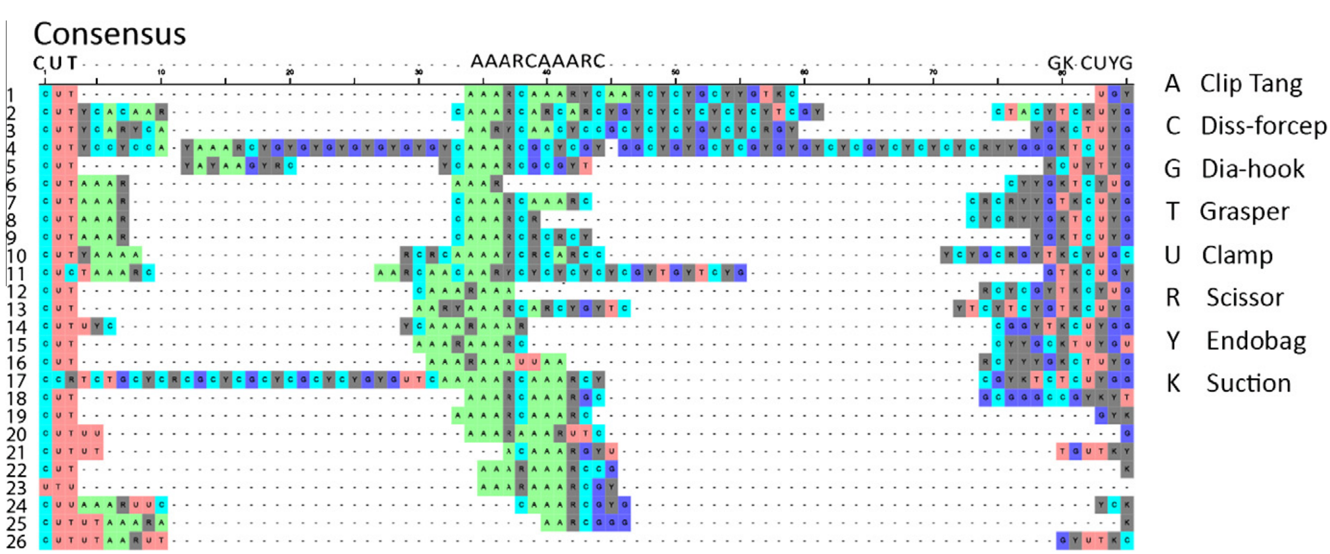

Fig. 4. Generating surgical consensus from surgical activity logs using multiple sequence alignment as described in steps 5 and 6 of Fig. 3.

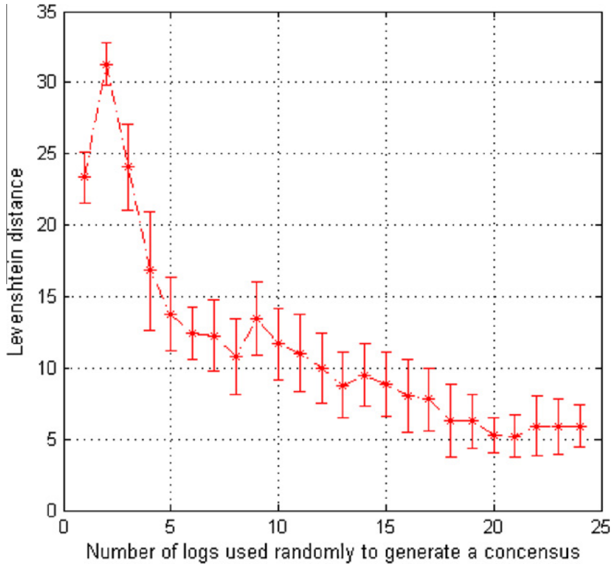

Fig. 5. Learning curve for deriving the consensus from a varying number of logs, the results are averaged over 20 randomly selected logs varying from 2 logs to 24.

inserted into the abdomen, the gallbladder is packed up into the bag and the bag extracted together with trocar (represented by Y). Finally, the surgical area is explored to detect and stop any bleeding (G). A drainage tube is inserted through a trocar hole and all instruments are removed.

To evaluate the stability of deriving a consensus from a varying number of activity logs, Fig. 5 shows the learning curve for deriving the consensus from 3 to 24 individual activity logs (represented on the *X*-axis). A random selection of logs is made for each number (3–24) to calculate the consensus and repeated 20 times in order to achieve more stable results. The *Y*-axis represents the Levenshtein distance (also known as the edit-distance) between the calculated consensus and the consensus derived in Fig. 4, which is based on all 26 activity logs. The Levenshtein distance metric is a commonly used metric to measure the similarity between the two strings. It is calculated as the number for deletions, insertions or substitutions required to transform one string to another. The learning curve starts flattening after 20 activity logs, which suggests that the consensus is stable and more activity logs would not help to reduce the distance much. Given that the calculated consensus from

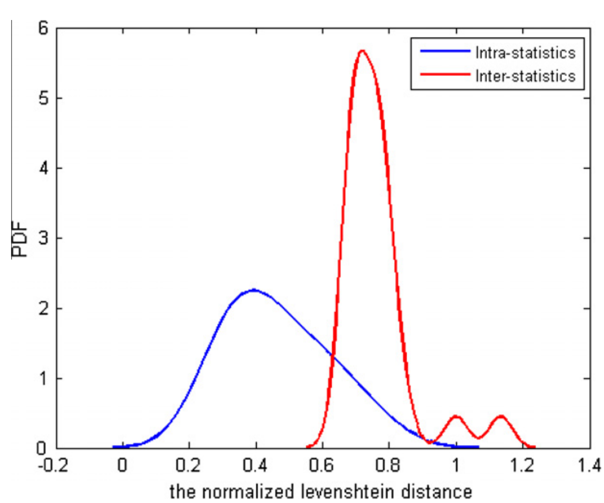

Fig. 6. Intra- and inter-statistics for discriminating between surgical and anomalous logs.

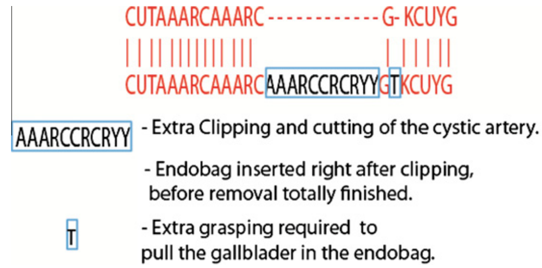

Fig. 7. Using global alignment to detect outliers.

the varying number of logs is compared with the consensus calculated from 26 activity logs, it is to be expected that the learning curve flattens as 26 is approached. As a result we cannot rule out that more activity logs would not significantly improve the consensus.

Fig. 6 shows the kernel density plot of the intra- and the inter-statistics performed to evaluate the discriminative performance of the derived consensus in discriminating between real-surgical logs and anomalous surgical logs. The intra-distances are calculated as the normalised levensthein distance between surgical logs and the consensus. The inter-distances are however the normalised levenshtein distance between random logs and the consensus. For a comparative analysis, the simulated logs are generated with the same length as the available surgical logs. The consensus clearly discriminate between the real- and random-logs, however there is a clear overlapping between the statistics. The overlapping area is due to the high variability in some of the surgical logs namely number 4 and 17 as illustrated in Fig. 3.

3.2. Outliers

Outliers are derived by calculating the differences between the consensus and an arbitrary activity log using a global pair-wise sequence alignment (Needleman–Wunsch) [7]. Needleman–Wunsch is a dynamic programming algorithm that assumes that the two sequences (consensus and activity log) are similar. It attempts to match them to each other from end to end and returns the alignment with an optimal global alignment score. The Needleman–Wunsch algorithm finds an alignment with the highest score where the score of an alignment is equal to the sum of the matches taken from the scoring matrix (i.e. identity matrix in our implementation (a matrix with ones in the diagonal)). For the treatment of gaps, the algorithm subtracts a penalty from the score for each gap space. Typically, the cost of extending a gap is many times lower than the cost for opening a gap.

As illustrated in, the algorithm inserts gaps (–) into the consensus in order to align it to the new activity log. The symbol | indicates activities that match exactly. Activities that do not match (no symbol |) together with the gaps (–) represent the deviations (i.e. outliers) from the treatment activity log from the consensus.

Note that all surgical procedures from Fig. 4 contain outliers (i.e. are different from the consensus), which is to be expected given the variable nature of surgical workflow. Those outliers are often simple variations in the execution of the surgical procedure, but they can also represent serious complications or errors. Fig. 7 provides a description of the outliers detected which helps to describe the surgery in more detail.

4. Conclusion and discussion

This paper has presented a new approach for deriving workflow consensus without prior knowledge using logs extracted from a clinical environment. We validated this approach for LAPCHOL surgical workflows. We have shown how multiple sequence alignment can be used to derive a workflow consensus which can then be used as the reference workflow for detecting outliers during surgery. The derived consensus includes the main steps of the LAPCHOL surgery as described in the best practices. We have also shown how outliers can be derived using pair-wise global alignment. Although it is important to monitor the compliance of surgical workflow with the consensus, it is also interesting to detect outliers. The outliers detected are deviations from the consensus and can represent both positive and negative deviations from standard surgical practice. The detected outliers can be used to enrich and extend medical protocols automatically in the case of good practices. Please note that time information about the outliers can be obtained from the original activity log.

For future work, we aim to start extensive data collection for the cost-classification of the outliers detected using this method. We therefore aim firstly to classify the outliers detected using a decision tree and then relate it to surgical cost measures (e.g. complications, post-operative outcomes, recovery time, etc.). In this way, the consensus (i.e. protocols) can be enriched by adding positively classified outliers as an alternative path in the consensus tree. Negative outliers, meanwhile, can be assigned a cost label that can be used in an objective clinical assessment of the procedure.

For future work, we also recommend using more data from other medical environments such as the ICU to generate the consensus workflow. We further recommend classifying the detected outliers using the cost-measures structured in a decision tree to allow a fully automatic description of outliers in those environments.

Acknowledgments

This research was supported by the Dutch grant organisation STW, Project No. 07320. The STW had no influence on the execution of this research. We would like to thank Dr. Armin Schneider for his comments on the outlier detection algorithm.

References

[1] van der Aalst WMP, van Dongen BF, Herbst J, Maruster L, Schimm G, Weijters A. Workflow mining: a survey of issues and approaches. Data Knowledge Eng 2003;47(2):237–67.
[2] Jagadeesh Chandra Bose R, Vander Aalst W. Trace alignment in process mining: opportunities for process diagnostics. Business Process Manage 2010:227–42.
[3] Rebuge Á, Ferreira DR. Business process analysis in healthcare environments: a methodology based on process mining. Inf Syst 2011.
[4] Bouarfa L, Jonker P, Dankelman J. Discovery of high-level tasks in the operating room. J Biomed Inform 2011;44(3):455–62.
[5] Huang Z, Lu X, Duan H. On mining clinical pathway patterns from medical behaviors. Artif Intell Med 2012.
[6] Lenz R, Peleg M, Reichert M. Healthcare process support: achievements, challenges, current research. Int J Knowledge-Based Organizations (IJKBO) 2012.
[7] Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 1970;48(3):443–53.
[8] Gribskov M, McLachlan AD, Eisenberg D. Profile analysis: detection of distantly related proteins. Proc Natl Acad Sci 1987;84(13):4355.
[9] Bruno G, Garza P. TOD: temporal outlier detection by using quasi-functional temporal dependencies. Data Knowledge Eng 2010;69(6):619–39.
[10] Bouarfa L, Akman O, Schneider A, Jonker PP, Dankelman J. In-vivo real-time tracking of surgical instruments in endoscopic video. Minimally Invasive Therapy Allied Technol 2011:1–6.
[11] OpenCV. <http://www.opencv.willowgarage.com/wiki/> [cited 2010 12-08-2010].
[12] Welch G, Bishop G. An introduction to the Kalman filter. Design 2001;7(1):1–16.